

University of Groningen

## A bed of ochre

Jelsma, Johan

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2000

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Jelsma, J. (2000). A bed of ochre: mortuary practices and social structure of a maritime archaic Indian society at Port au Choix, Newfoundland. s.n.

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

---

## 5 Statistical Analyses

*"Intuitions of obscure origin are splendid, in fact, probably essential, for thinking up really good new ideas. But intuition is not a satisfactory basis for assessing the fit between ideas and evidence."*

(Cowgill 1977:352).

In order to detect which part of the archaeologically observed variation in mortuary attributes reflects the original burial practices of a prehistoric society, one needs to take into account the possibility of biases in the data, such as by post-depositional processes, excavation and publication errors, as well as chance. One must look for meaningful patterns in the mortuary data rather than for absolute distinctions. To trace such patterns statistical analyses are essential. A description of the statistical analyses used in this research is presented below. For a general introduction to the use of statistics in archaeology the reader is referred to Shennan (1988). For a discussion on the mis-use of statistics in archaeological research the work of Thomas (1978) is to be recommended.

### **Uni and Bivariate Statistics**

By means of statistical probability analysis it is possible to calculate how likely it is that some observed variation in the data was either caused by a real difference, or by chance. Before conducting a statistical test, one needs to decide at which point, or significance-level, one should consider the observed variation either as a real difference, or exclusively as due to chance. Often used significance levels, or  $\alpha$ , by which statistical test results can be

interpreted are 0.01 or 0.05. This means that a particular variation in the data is considered to be statistically significant, or real, if the probability that this variation was caused by chance, is one/five percent or less. The probability  $\alpha$  is the chance that the null hypothesis ( $H_0$ ), which predicts randomness, is rejected unjustly in favour of the alternative hypothesis  $H_1$ , which predicts a difference (Siegel & Castellan 1988:7).

So the researcher needs to make a choice between  $H_0$  and  $H_1$ . In doing this he or she can make two errors. The first is to reject  $H_0$  (which predicts randomness) while in fact that hypothesis is true. This is called a Type I error. A Type II error occurs when  $H_0$  is not rejected while it is false, and instead  $H_1$  (which predicts a difference) is true. The probability of making a Type I error is  $\alpha$ . For a Type II error this probability is  $\beta$ , which is a measure of the power of a statistical test. If the results of probability tests are below a certain  $\alpha$  value, they are considered statistically significant. The higher  $\alpha$  value (significance level) we choose, the higher the probability of the occurrence of Type I errors becomes. Or in other words; the more likely it becomes that test results we call statistically significant are caused by chance. If a very low  $\alpha$  value is chosen as the significance level, decreasing the probability of Type I errors, the likelihood of Type II errors ( $\beta$ ) increases. So it becomes more likely that probability test results are not called statistically significant while they do reflect real differences. In order to balance the risks of Type I and Type II errors, to be critical but not throw the baby out with the bath-water,  $\alpha$  and  $\beta$  need to be equalised (Siegel & Castellan 1988:8-10). The study of Port au Choix-3 locus II involves the analysis of one single site, which was excavated in a consistent manner by one excavation team in a short period of time. Therefore an  $\alpha$  of 0.05 was considered to be the best balance for the analysis of the archaeological data from this cemetery. In cases where two or more sites are compared statistically an  $\alpha$  of 0.01 is more appropriate (Newell & Dekin 1978). In this publication statistically significant probabilities ( $p$ ) at an  $\alpha$  of 0.05 are printed in **underlined bold face**, probability results between 0.05 and 0.10 are presented in **bold face**.

The rejection of  $H_0$  is dependent upon the nature of the alternative hypothesis  $H_1$ . In a statistical experiment where  $H_1$  only predicts a difference between the distributions of two sampled populations, or between that of one sampled population and a theoretical distribution, and not the direction of that difference, a two-tailed test must be used. If  $H_1$  predicts that direction a one-tailed test must be used (Siegel & Castellan 1988:14). For instance in the statistical comparison of female and male canine (dental) sizes, one must use a one-tailed test because on the basis of biological data one can predict that male canines are larger than female ones. Before a statistical test is carried out, one has to decide whether a one-tailed or two-tailed probability is appropriate. The two-tailed probability is twice as high as the one-tailed probability.



**Figure 5.1.** Gouge (attribute 101) from burial C 50A. Length of the object is approximately 15 cm. Drawing by Katherine Scott. (EeBi-2:1537).

The significance tests used here are all of the non-parametric or distribution free category. In contrast with parametric tests, such as the Student's-t test, non-parametric tests make less stringent assumptions about the distributions of the sampling populations from which our data originate. The use of parametric tests involves the assumption that our samples come from a normally distributed population (Siegel & Castellan 1988:3). Such distributions can be fully and accurately characterised by their mean and standard deviation (*ibid.*:13). Archaeological variables are very often not normally distributed, and therefore do not meet the requirements of most parametric tests. Therefore non-parametric tests are more suitable for the analysis of archaeological samples. Other advantages of non-parametric tests are their applicability to very small sample sizes, and the possibility to analyse data measured on a nominal scale (*ibid.*:35). In contrast with ordinal data, nominal data cannot be ranked in any way. Two other scales exist.

These are the interval and ratio scales. In contrast to an ordinal scale, the distances between the numbers on the interval scale are known (e.g. the Celsius temperature scale). In the ratio scale those distances are also known, but the ratio scale has a true zero point. For example weight is measured on a ratio scale.

The statistical analysis of one-sample cases is a comparison between the frequency distribution observed in the sample, and the frequency distribution expected under  $H_0$ . For the statistical analysis of one-sample cases, the binomial test, and the Chi-square goodness-of-fit test were used for nominal data. If a sampling population consists of only two classes (e.g. females and males), a binary population, it can be analysed with the binomial test. This test calculates the probability that a sample originates from a population with a certain, usually equal, distribution of the two classes (Siegel & Castellan 1988:38). If the cases in a sampling population are distributed over two or more (nominal) classes, the Chi-square goodness-of-fit test can be used. This test calculates whether there is a statistically significant difference between the frequency distribution of a sample, and the frequency distribution that is expected under the null hypothesis ( $H_0$ ) (*ibid.*:45).

If the data in a one-sample case are continuous and measured on an ordinal scale they can be analysed by the Kolmogorov-Smirnov single-sample test which calculates the degree of correspondence between an observed cumulative frequency distribution and the cumulative frequency distribution expected under the null hypothesis ( $H_0$ ) (*ibid.*:51).

In the cases that frequency distributions of two independent samples are available, these can be compared to each other. In this way it can be assessed whether the two underlying sampling populations differ significantly. For discrete nominal data distributed into two classes, two statistical tests are available; the Fisher exact probability test used for small sample sizes, and the Chi-square test for 2x2 tables applicable to larger samples. These tests calculate whether there are any statistically significant differences between the frequency distributions of two independent samples, whose cases belong to one of two mutually exclusive classes (Siegel & Castellan 1988:103,111). An example is the statistical comparison between a group of female and a group of male burials (the two independent samples), with respect to the presence or absence (the two mutually exclusive classes) of a certain grave good. The Fisher exact probabilities presented in this publication are all two-tailed probabilities.

In the case of two samples with more than two classes the Chi-square test for 2 x k tables or, in case of low frequencies, the multiple contingency test can be used (Verbeek *et al.* 1983). The last was conducted using the FISHER 3.0 programme (Verbeek & Kroonenberg 1990). Also for the executed FISHER multiple contingency tests, all probabilities are two-tailed.

If we are dealing with continuous data, other tests are suitable for the statistical analysis of two independent samples. Firstly the Mann-Whitney U test can be used (Siegel 1956:116), which is appropriate for uni-modal frequency distributions. In case of bimodal or multi-modal distributions the Kolmogorov-Smirnov two-sample test can be applied. This test calculates whether the cumulative frequency distributions of the two samples are significantly different from each other (Siegel & Castellan 1988:51).

In addition to the uni-variate significance testing described above, bivariate tests were conducted as well. The aim of this kind of statistical analysis is to establish to what degree two preferably interval, but at least ordinal variables are associated or correlated with each other. As a measure of correlation, the Spearman rank order correlation coefficient ( $r_s$ ) was used. This test calculates the differences between the rank order of two variables of each object to determine the degree of positive or negative correlation between those two variables (Siegel & Castellan 1988:235). The uni- and bivariate analyses were conducted using the SPASPAC programme package version 5.1 (Newell 1995) and the FISHER 3.0 programme mentioned earlier.

## **Multivariate Statistics**

Multivariate techniques are valuable tools for the exploration of data bases with more than two variables. These techniques detect the presence of sub-groups and/or predominant third dimensions (trends) in such data bases, but most such techniques cannot determine whether these sub-groups are significantly different from each other. In order to screen such sub-groups for significant differences one needs to conduct a number of uni- and/or bivariate probability tests. Multivariate techniques are not often enough applied in archaeology. For a compilation the reader is referred to the work of Baxter (1994). In the statistical analysis of the Port au Choix-3 locus II data three groups of multivariate techniques were used. These are numerical ordination and seriation (1), principal component analysis and/or correspondence analysis (2), and cluster analysis (3).

For the analysis of presence/absence data numerical ordination and seriation are useful techniques. They were first introduced into archaeology by Petrie (1899). Ordination is a series of two-dimensional relocations of cases, such as graves or sites, and their attributes in a presence/absence matrix, aiming at the rendering of visually observable groupings in the data. In ordination the cases, which are presented as rows in the matrix, are ranked in such a way that the presence dots in each attribute column are as close together as possible. The degree of compactness of this ranking is indicated by the measure of concentration, which is the number of presence dots divided by the total number of cells enclosed in a matrix. The higher the

measure of concentration, the better the cases are grouped according to their similarities and associations.

Subsequently in an ordination the attributes are ranked in such a manner that the presence dots of the matrix are contiguously and symmetrically located around and along the X/Y diagonal. In this way, groups of cases with their characteristic attributes can be separated from other groups having their own distinct attributes (e.g. Figure 7.11.). One of the underlying assumptions of ordination is that the role of third dimensional processes, usually chronology, is limited.

In order to investigate the influence of such third dimensional processes a seriation should be conducted. In a seriation the attributes are grouped in a sequential order. The cases are ordered according to their resemblances in attribute compositions. This order is determined by the first occurrence of an attribute going through the sequence of ranked cases. (e.g. Figure 7.12.). If the patterning in the data is mainly determined by a third dimension and the cases are ordered according to that dimension, the highest presence dots of each column are located along an approximately straight line; the seriation line (Rozoy *et al.* 1991). Examples of such third dimensions are time, a spatial cline, status hierarchy and seasonal sequence. Ordination and seriation were conducted using the ROZOY Numerical Ordination and Seriation Program Package (*ibid.*).

The second approach to the detection of patterning in multivariate data is the application of principal component analysis and correspondence analysis (Greenacre 1984). Correspondence analysis, which is in fact a variant of principal component analysis, is suitable for the exploration of discrete nominal data. For the treatment of continuous data principal component analysis is more appropriate (Baxter 1994:100). Both techniques aim at the replacement of the total, and large number of variables in a certain data base, by a smaller number of new and uncorrelated variables. These techniques can only be useful if there is a high level of inter-correlation among the original variables (*ibid.*:50). The correlated, original, variables are summarised to components, or axes, which are linear equations related to these variables. The first component or axis is along the greatest spread in the multi-dimensional point scatter. It therefore represents the highest proportion of variation present in the data and is the most explanatory dimension. The second axis represents the next highest, and so on. If the original variables are highly correlated, the first few components or axes can summarise most of the variation present in the data set (Doran & Hodson 1975). Major patterning in the data can be visualised by making a series of two dimensional scatterplots of those first few components or axes. However, if the bulk of the variation in a particular data base cannot be covered by the first few components or axes, the visualisation of this variation by means of scatterplots can be very difficult. In such cases it is possible to conduct cluster analysis on the new sets of uncorrelated variables

which resulted from principal component or correspondence analysis (Baxter 1994:169). In the statistical analysis of the Port au Choix data, principal component analysis and correspondence analysis were conducted using the programme MVSP - Multi Variate Statistical Package 2.2 (Kovach 1995). These algorithms were only used to obtain new sets of less correlated variables. The principal component and correspondence analysis scores were further analysed with K-means divisive cluster analysis.

Either applied in combination with principal component analysis/ correspondence analysis, or on the data directly, cluster analysis is a very useful range of techniques for the exploration of multivariate data. Two basic steps can be distinguished in clustering procedures. Firstly the difference or similarity between each pair of cases is established according to a chosen distance or similarity measure, such as the Euclidean or squared Euclidean distance. Secondly the cases are combined into groups following a chosen algorithm. Three main kinds of algorithms exist. These are hierarchical monothetic divisive, hierarchical agglomerative, and K-means clustering. Hierarchical monothetic divisive procedures were not used in this study because outliers and rare attributes often led to spurious divisions of the data sets (Baxter 1994:150). Hierarchical agglomerative clustering procedures start with each object as a separate cluster. At each step in the procedure the two clusters which are most alike are combined. The main advantage of such procedures is that the successive combinations of the units can be visualised in a dendrogram. A disadvantage is the fact that when an object is designated to a certain cluster, its position is fixed even though in a later phase of the clustering procedure its optimal place may be in another cluster (Hodson 1969; Baxter 1994). However, some clustering programs, such as CLUSTAN (Wishart 1987), are equipped with a 'relocate' option which offers the possibility to move classified cases to other, more suitable, clusters. Another approach, which is optional in MVSP (Kovach 1995), is to randomise the input of data and to conduct multiple runs of the same coefficient/clustering method algorithm.

The cluster analyses used in this research were conducted by means of K-means divisive clustering procedures. This type of clustering is not widely used in archaeology, although it is considered to be more appropriate for most archaeological data than are the hierarchical procedures (e.g. Baxter 1994:159). Possible reasons for this are formulated by Wright;

*"The prevalent use of hierarchical models probably reflects habit and the availability of programs, not the usefulness of the hierarchical model. Furthermore a dendrogram is handsome, whereas K-means assignment to groups has no visual appeal."*

(Wright 1992:99).

For the analyses of the Port au Choix study the MV-NUTSHELL programme K-means (Wright 1994) was used. On a particular data set the MV-NUTSHELL programme conducts, as a maximum, nine classifications in



which it calculates the best fit of the cases into two to 10 clusters. The K-means clustering procedure consists of the allocation of the cases between the clusters, until the optimal grouping has been reached. The programme also gives a graphic rendition of the K-means results, which facilitates the decision on the optimal number of clusters. This decision is based on the D values which are calculated for each number of clusters (Spath 1988 cited in Wright 1992:100). Those D values are plotted in the graph which can be used to determine the number of groups present in the data. A smooth slope of the D value curve indicates that the data should not be subdivided at all. On the other hand, the existence of one or more 'elbows' in the slope indicates where the data are optimally subdivided. The location of the most acute elbow determines the number of groups into which the cases should be classified (Wright 1992:100, 1994:3).

Principal component analysis, correspondence analysis, and K-means cluster analysis were all conducted in Q-mode, in which the cases are classified, and, in some instances, in R-mode, which classifies the attributes.

This completes the description of statistical techniques used in the Port au Choix-3 locus II study. A list of computer programmes used in the Port au Choix research is presented in Appendix 5.1. In Chapter 6 the methods and techniques concerning the analysis of the human skeletons are discussed.